Executive Steering Committee
For A.C.E. Policy II
(ESCAP II)
Report 13
October 9, 2001

ESCAP II: Effect of Excluding Reinstated Census People from the A.C.E. Person Process

David A. Raglin
Planning, Research, and
Evaluation Division

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

EXECUTIVE SUMMARY

What was the effect on the results of the A.C.E. person process by excluding reinstated census people?

The effect of excluding reinstated census people from the A.C.E. person process was minimal. It is estimated that the A.C.E. coverage rate was overestimated by only 0.034 to 0.082 percentage points, so the dual system estimate (DSE) was underestimated by about the same percentage.

This result confirmed the assumption previously made by Hogan in "Data and Analysis to Inform the ESCAP Report", DSSD Census 2000 Procedures and Operations Memorandum Series B-1*, dated March 1, 2001, that the effect of the reinstated people on the DSEs would be small.

Definition of "reinstated census people"

Some people in the census were not included in the A.C.E. person process. These were people that were initially deleted from the census but later reinstated into the census after the A.C.E. person process had begun. They were initially deleted because they were suspected of being duplicates of other census housing units.

Relationship of reinstated census people to the DSEs

In "Accuracy and Coverage Evaluation Survey: Effect of Excluding 'Late Census Adds'", Internal Census Bureau memorandum, DSSD Census 2000 Procedures and Operations Memorandum Series # Q-43, dated March 22, 2001, Hogan showed if the reinstated people were a small percentage of the correct enumerations in the census or their A.C.E. coverage rate (percentage of census correct enumerations that were found in A.C.E.) was similar to the A.C.E. coverage rate for census people included in A.C.E., there would be a minimal effect on the DSEs.

Methodology

We clerically matched the reinstated people collected in A.C.E. and census in evaluation clusters (a 1/5th sample of A.C.E. clusters), attempting to mimic as best as possible what would have happened had they been among the census people in the production matching operations.

We determined in the matching operation if the people were:

- Matched to an A.C.E. person who was living in the cluster on Census Day, and therefore were correct enumerations
- Duplicated census people, and therefore were erroneous enumerations
- Found in neither the A.C.E. nor the census. These people had undetermined enumeration status.

We expected that a substantial number of the reinstated people would be duplicates of census people already included in the A.C.E. person process because person duplication was considered as part of the deletion and reinstatement process.

In production, the people not found in the Census or A.C.E. would have been included in the Person Followup (PFU). We could not follow these people up, so we do not know if they are correct or erroneous enumerations. We found that 22.1 percent of the reinstated people were in that category. We assumed several different correct enumeration rates for those people in a sensitivity analysis to provide a range of the effect on the A.C.E. coverage rate for these people and for the effect on the overall A.C.E. coverage rate.

The effect of this operation on the results

We found:

- 558,448 matched correct enumerations (25.4 percent of the 2,198,492 total)
- 486,626 people with undetermined enumeration status–that we could not find in A.C.E. or Census (22.1 percent)
- 1,153,418 erroneous enumerations--almost all duplicates of Census people (52.5 percent)

Hogan showed that reinstated people who turned out to be erroneously enumerated did not have an effect on the DSE, so only the matched and undetermined people need to be taken into account to determine the effect on the DSE.

The sensitivity analysis based on varying correct enumeration rates for the people with undetermined enumeration status indicated that the A.C.E. coverage rate was overestimated by between 0.034 and 0.082 percentage points by excluding the reinstated people from the A.C.E. person process. Clusters with P-sample large block subsampling, relist clusters, and housing units with an initial housing unit match code indicating geocoding error or an erroneous enumerated housing unit were excluded from this analysis.

Also, the distribution of correct, erroneous, and undetermined enumeration status and the distribution of the estimated number of reinstated people is similar for large poststratification demographic groups, so there is no reason to believe that there might be a differential effect on any particular poststrata.

1. BACKGROUND

1.1 The Summary of the Conclusion

This report answers the question: what was the effect on the A.C.E. process of not including people that were initially dropped from the Census and later reinstated–called reinstated people. I show the change in correct enumerations and erroneous enumerations, as well as identify people for which I cannot make a determination of their enumeration status.

Excluding the reinstated census people from the A.C.E. person process had a minimal effect on the A.C.E. results. This confirms the assumption made by Hogan (2001a) that there was no reason to believe that excluding the reinstated people would affect the results of the A.C.E.

1.2 Census people not included in the A.C.E. person process

Some people in the Census were not included in the A.C.E. operations:

- During the processing of Census 2000, the Census Bureau reinstated into the census certain housing units, and the people reported in those housing units, that had been suspected of being duplicates of other census housing units.
- However, the reinstated units and their people were not included in the census-to-A.C.E. person matching process, and therefore they were not included in the DSEs.

1.3 How the reinstated people relate to DSEs

Table 1 shows the standard method of classifying people for the DSEs. The A.C.E. process provides cells n_{11}, n_{21}, and n_{12} (and therefore cells n_{+1} and n_{1+}) and estimates cell n_{22} and therefore the rest of the total cells.

Table 1: Classifying Census and A.C.E. People for DSE

	A.C.E. Persons		
Census Persons	Found in Census	Not found in Census	**Total Correctly in A.C.E.**
Found in A.C.E.	n_{11}	n_{12}	$\mathbf{n_{1+}}$
Not found in A.C.E.	n_{21}	n_{22}	$\mathbf{n_{2+}}$
Total Correct Enumerations in Census	$\mathbf{n_{+1}}$	$\mathbf{n_{+2}}$	$\mathbf{n_{++}}$

Note the numbers in this table are weighted and only include census and A.C.E. people in A.C.E. clusters. This table is formed individually for each A.C.E. poststrata (based on age, sex, geography, and race/Hispanic origin).

Cell n_{+1} does not represent all people collected in the census. The A.C.E. process determined that some census people were erroneous enumerations, and they do not appear in this table. Table 2 is a summary of how A.C.E. categorizes all of the people in the census:

Table 2: Classification of Census People for DSE

All People Found in the Census (called n_{+1t})	
Found in A.C.E (cell n_{11} from Table 1)	Total Correct Enumerations in Census (cell n_{+1} from Table 1)
Not found in A.C.E. (cell n_{21} from Table 1)	
Erroneously Enumerations in Census (called n_{+1e})	

In production, the reinstated people were not included in the numbers used to produce the counts in the tables above since they were not part of the A.C.E. person process. If they had been in the process, they would have been classified into one of the three groups above.

How much of an effect does excluding the reinstates have on the final DSEs? According to Hogan (2001b), two factors determine the magnitude of the effect in a poststratum:

- The proportion of census correct enumerations that match A.C.E. people (n_{11} / n_{+1}) -- the A.C.E. coverage rate -- for the reinstates compared to other census people
- The number of correctly enumerated reinstated people relative to the number of other census people

Therefore, this report focuses on determining how many reinstated persons fall into each cell.

2. METHODS

2.1 Clerically matching the reinstated people

We conducted a clerical matching operation at the National Processing Center (NPC) to match the reinstated people to the people previously available from the A.C.E. and the census. The philosophy of the matching operation was to simulate as best as possible what would have happened if the reinstated people had been among the census people available to match from the start.

We matched reinstated people in the evaluation clusters, a 1 /5th sample of the A.C.E. clusters. These are the same clusters used in such evaluations as the Analysis of Measurement Error, the Matching Error Study, and the Total Error Analysis. Reinstated people in those clusters and people in surrounding blocks to A.C.E. Targeted Extended Search clusters will be included in this operation. Including people in all parts of the evaluation clusters and people in surrounding blocks of A.C.E. Targeted Extended Search (TES) clusters, there were 9,813 reinstated people in 3,678 housing units in 879 clusters.

Production clerical person matching was done using a computer-based system called Per MaRCS (Person Matching and Review Coding System). We used a hybrid of Per MaRCS and custom printouts of the reinstated people to do this matching. The matchers had the same information for the reinstated people as they had for the census people in production. The PerMaRCS system was used in read-only mode to see the results of production matching.

2.2 Tabulation of the results

Ideally, we would be able to determine which DSE category from Table 2 each reinstated person should fit into. Table 3 shows how the results from matching map into the DSE categories:

Table 3: Obtaining classifications for DSEs from the matching results

Matching Result	Classification for DSEs
Matched an A.C.E. person	Found in A.C.E., correctly enumerated in the census (n_{11})
Duplicate of a census Person already in the system	Erroneous enumeration (n_{+1e})
Neither found in A.C.E. or census	Cannot determine, could be either: • Not found in A.C.E., correctly enum. in census (n_{21}) • Erroneous enumeration (n_{+1e})

We cannot determine if a reinstated person who is not a match to an A.C.E. person or a duplicate of a census person was a correct enumeration missed in A.C.E. (cell n_{21}) or an erroneous enumeration (cell n_{+1e}). The production process allowed us to make a distinction for the "Neither found in A.C.E. or census" group using the Person Followup survey, which collected residence information about unmatched census people. We cannot do a similar followup to collect quality data about Census Day residence over a year after Census Day.

There are also some unusual situations that happened occasionally, as they did in production matching. Sometimes an E-sample person is the same person as a person in a part of the cluster that was subsampled out, which affects the correct enumeration probability. Sometimes the E-sample person does not have sufficient information for matching (name and two characteristics) or is in a housing unit that was incorrectly geocoded in the cluster and is therefore considered an erroneous enumeration for that reason. Therefore, the tables in this report are actually based on the change in each type of enumeration, not just the weighted estimate of the reinstated people.

3. LIMITATIONS

As was mentioned in the previous section, we were unable to conduct a followup operation to determine if people not found in the A.C.E. or in the census were in actuality correct enumerations or not. We have to instead make assumptions about what percentage of these people were in actuality nonmatched correct enumerations as opposed to erroneous enumerations.

We attempted to identify E-sample housing units in a manner similar to production. As implemented, the procedure was different than the production procedure for clusters involved in large block subsampling but we were able to calculate weights and estimates given the E-sample selection process that was used.

4. RESULTS

4.1 Change in the Number of People by Enumeration Status

Table 4 shows the change in correct, undetermined, and erroneous with the addition of the reinstated people to the census. Standard errors are in parenthesis.

Table 4: Change in Each Type of Enumeration Status Based on the Reinstated People

Enumeration Status	Estimate	Percent
Matched to A.C.E. - correctly enumerated (n_{11})	558,448 (92,194)	25.4%
Not found in A.C.E. or census – undetermined enumeration status	486,626 (64,836)	22.1%
Erroneously enumerated (mostly duplicates of census people)	1,153,418 (112,350)	52.5%
Total	2,198,492 (165,939)	100.0%

Slightly more than half of the change in the number of E-sample people was due to erroneously enumerated people. Most of those were people duplicated to census people (about 97 percent of the erroneous enumerations); a few were people with insufficient information for matching and geocoding errors. The high duplication rate between reinstated census people and census people included in the A.C.E. person process is not a surprise because person duplication was considered during the initial deletion and reinstatement process (Nash, 2000).

We could not determine the enumeration status for 486,626 people. These people did not match A.C.E. people nor were they duplicates of census people. In production, they would have gone to followup and many would have been classified as correct or erroneous enumerations.

4.2 Effect of the Reinstates on the Overall A.C.E. Coverage Rate

What was the effect on the A.C.E. coverage rate by excluding the reinstated people? We know the matches are correct enumeration matches and some of the undetermined people are nonmatched correct enumerations. We can do a sensitivity analysis varying the percentage of the undetermined people who are correct enumerations and measure the effect on the overall A.C.E. coverage rate.

The A.C.E. process was designed for A.C.E. to measure Census coverage, not the other way around. When Beaghen (2001) measured A.C.E. coverage, he excluded clusters with P-sample large block subsampling because there could be Census nonmatches because of the A.C.E. and Census samples not overlapping. He also excluded people in housing units with an initial housing unit match code indicating geocoding error or an erroneously enumerated housing unit, and relist clusters because the census non-matched people could in actuality be outside the

cluster.

Of the 558,448 matches in Table 4, 384,520 are left after those exclusions, 255,613 of the 486,626 undetermined people, and 615,889 of the 1,153,418 erroneous enumerations are left. Note that a larger percentage of undetermined people are excluded than matches–that indicated that the effects that Beaghen was concerned about are valid for the reinstated people.

Table 5 shows the results of the sensitivity analysis based on different correct enumeration rates of the undetermined people – 38, 50, and 70 percent. The excluded people described before are not included. The 38 percent figure was chosen because it is the correct enumeration rate of the resolved people included in the study (384,520 / (384,520 + 615,889)).

Table 5: The Effect of the Reinstates on the Overall A.C.E. Coverage Rate

Enumeration Status	Without Reinstated People	If this percentage of the undetermined people were correct enumerations		
		38%	**50%**	**70%**
Matched – correct enumerations (CEs) (n_{11})	-------	384,520 (81,245)	384,520 (81,245)	384,520 (81,245)
Nonmatched - assumed CEs (n_{21}), based on the % of the 255,613 undetermined people	-------	97,133 (16,956)	127,806 (22,310)	178,929 (31,234)
Total CEs (n_{+1}), equal to Matched + Nonmatched	-------	481,653 (84,198)	512,327 (85,808)	563,449 (89,144)
A.C.E, coverage rate: % matched among reinstated people, equal to Matched / Total	-------	79.8% (4.5)	75.1% (5.2)	68.2% (6.0)
A.C.E. coverage rate: % matched among all people	90.732% (0.291)	90.698% (0.290)	90.680% (0.291)	90.650% (0.291)
Change in A.C.E. coverage rate due to inclusion of the reinstated people	-------	**-0.034% (0.011)**	**-0.052% (0.014)**	**-0.082% (0.019)**

Note that the change in the overall A.C.E. coverage rate by including the reinstated people is small – dropping only 0.034 to 0.082 percentage points in the different scenarios presented in Table 5 – which would cause, all else being equal, a similarly small increase in the DSE. This is because, based on the factors discussed in Section 1.2 that Hogan (2001b) developed:

- There are not many reinstated people in the census once erroneously enumerated people are removed – 0.37 percent at the most based on the data in Table 5, and
- The match rate of the reinstates is not radically different than the match rate of census people overall, given the small number of people involved. The best-case A.C.E. coverage rate for these people is almost 80 percent and the worst-case is still almost 70 percent.

4.3 Reinstate Estimates for Poststrata Demographic Variables

Even though the reinstated cases have no effect at the national level, there is the possibility that they disproportionally affect certain types of people or poststrata. Table 6 gives the correct, undetermined, and erroneous enumeration estimates for the age/sex, tenure, and domain poststrata variables. Standard errors are in parenthesis. The correct, undetermined, and erroneous percentages (%Cor, %Und, and %Err) add to 100 percent in each row. The % reinstated (% Rein) and % of population (% Pop) add to 100 percent in each column.

Table 6: Changes in Enumeration Status by Poststrata Variables

Age/Sex	Correct	Undet.	Erron.	Total	%Cor	%Und	%Err	% Rein	% Pop
0-18 All	160,006 (40,672)	141,450 (30,935)	387,182 (51,383)	688,639 (73,859)	23.2%	20.5%	56.2%	31.3%	25.9%
18-29 Male	47,022 (12,464)	48,429 (12,276)	67,821 (12,952)	163,272 (22,004)	28.8%	29.7%	41.5%	7.4%	7.5%
18-29 Female	53,887 (12,655)	61,942 (15,105)	74,429 (13,887)	190,258 (24,378)	28.3%	32.6%	39.1%	8.7%	7.7%
30-49 Male	63,956 (16,441)	68,012 (15,523)	162,956 (21.326)	294,923 (31,725)	21.7%	23.1%	55.3%	13.4%	15.2%
30-49 Female	83,376 (19,659)	58,348 (11,472)	187,258 (23,317)	328,983 (33,181)	25.3%	17.7%	56.9%	15.0%	15.8%
50+ Male	65,561 (14,020)	50,187 (10,982)	130,802 (20,040)	246,550 (27,449)	26.6%	20.4%	53.1%	11.2%	12.7%
50+ Female	84,640 (19,069)	58,258 (12,731)	142,969 (19,799)	285,867 (30,915)	29.6%	20.4%	50.0%	13.0%	15.2%
Total From Table 4	558,448 (92,194)	486,626 (64,836)	1,153,418 (112,350)	2,198,492 (165,939)	25.4%	22.1%	52.5%	100.0%	100.0%

Tenure	Correct	Undet.	Erron.	Total	%Cor	%Und	%Err	% Rein	% Pop
Owner	343,829 (74,047)	250,355 (43,804)	761,251 (90,976)	1,355,435 (131,452)	25.4%	18.5%	56.2%	61.7%	70.3%
Renter	214,619 (40,510)	236,271 (48,309)	392,167 (62,457)	843,058 (91,449)	25.5%	28.0%	46.5%	38.3%	29.7%
Total From Table 4	558,448 (92,194)	486,626 (64,836)	1,153,418 (112,350)	2,198,492 (165,939)	25.4%	22.1%	52.5%	100.0%	100.0%

Domain	Correct	Undet.	Erron.	Total	%Cor	%Und	%Err	% Rein	% Pop
Am Ind On Res	3,923 (2,168)	488 (400)	1,374 (1,001)	5,785 (2,451)	67.8%	8.4%	23.8%	0.3%	0.2%
Am Ind Off Res	2,318 (1,445)	7,136 (6,513)	10,280 (6,322)	19,733 (9,462)	11.7%	36.2%	52.1%	0.9%	0.6%
Hispanic	64,977 (26,912)	53,466 (17,706)	194,880 (43,087)	313,323 (61,154)	20.7%	17.1%	62.2%	14.3%	12.1%
Black	90,789 (33,262)	68,290 (19,260)	184,074 (39,736)	343,153 (57,897)	26.5%	19.9%	53.6%	15.6%	11.6%
Hawaii/ Pac Isl	10,680 (9,503)	0 (0)	0 (0)	10,680 (9,503)	100.0%	0.0%	0.0%	0.5%	0.2%
Asian	23,947 (25,452)	12,930 (6,295)	32,422 (13,353)	69,299 (32,474)	34.6%	18.7%	46.8%	3.2%	3.3%
White/ Other	361,813 (57,973)	344,316 (55,884)	730,389 (92,081)	1,436,518 (124,556)	25.2%	24.0%	50.8%	65.3%	72.2%
Total From Table 4	558,448 (92,194)	486,626 (64,836)	1,153,418 (112,350)	2,198,492 (165,939)	25.4%	22.1%	52.5%	100.0%	100.0%

The %Cor, %Und, and %Err columns show that the distribution of correct, erroneous, and undetermined enumeration status for reinstated people and the distribution of the estimated number of reinstated people (shown in the % Rein column compared to the % Pop column) is similar for large poststratification demographic groups, especially when the standard errors of the point estimates are taken into account.

5. CONCLUSIONS

While it certainly would have been ideal to have included all census people in the A.C.E. person process, that was not possible. Hogan (2001b) showed that if there were relatively few correct enumerations added or the match rate of these people was relatively similar to the overall match rate, the effect on the dual system estimates would be minimal. This report found both of those conditions to be true, and therefore the change in the match rate after taking into consideration the reinstated people is very small.

References

Beaghen, Michael (2001). "Accuracy and Coverage Evaluation: TES Balancing". Internal Census Bureau memorandum, February 7, 2001.

Hogan, Howard (2001a). "Data and Analysis to Inform the ESCAP Report", Internal Census Bureau memorandum, DSSD Census 2000 Procedures and Operations Memorandum Series #B-1*, March 1, 2001.

Hogan, Howard (2001b). "Accuracy and Coverage Evaluation Survey: Effect of Excluding 'Late Census Adds'". Internal Census Bureau memorandum, DSSD Census 2000 Procedures and Operations Memorandum Series # Q-43, March 22, 2001.

Nash, Fay (2000). "Overview of the Duplicate Housing Unit Operations". Internal Census Bureau Document dated November 7, 2000.

cc: ESCAP members
K. Styles (DMD)
F. Nash
M. Urrutia
J. Stuart
S. Brady
S. Walker
K. Porter (DIR)
D. Hubble (PRED)
R. Petroni
D. Raglin